DOCUMENT RESUME

ED 056 993 SP 005 396

AUTHOR English, Fenwick W.; And Others
TITLE Evaluating the Effects of Implementing a Differentiated Teaching Staff: Problems and Issues. A Tentative Position Paper for Use in Project Evaluation.
INSTITUTION Mesa Public Schools, Ariz.
PUB DATE Nov 71
NOTE 20p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Differentiated Staffs; Elementary Schools; *Evaluation Methods; Junior High Schools; *Pilot Projects; *Program Evaluation; *Research Design; School Personnel

ABSTRACT

Since the inception of the Arizona-Mesa Differentiated Staffing Project in June 1970, the project staff was charged with the responsibility of evaluating the effects of the changes brought about as a result of implementation in three pilot schools. It was found that experimental-design research in an ongoing social system was extremely difficult because of inability to control internal and external variables. After reviewing various methods used to evaluate social innovations, the staff decided on a combination of the case-study technique advocated by Weiss and Rein and the quasi-experimental design advocated by Campbell. Factors jeopardizing internal and external validity were identified and tentative research designs were outlined for five questions. The questions involved comparing project and non-project schools in relation to student achievement, school climate, staff attitudes, teacher ability to perform tasks specified in the original proposal, and degree of actual staff differentiation. Each research design specifies a hypothesis, data-gathering instruments, dependent and independent variables, and statistical treatment. (RT)

ARIZONA-MESA DIFFERENTIATED STAFFING CONSORTIUM
U.S. Office of Education, PL 80-35
MESA PUBLIC SCHOOLS

George N. Smith
Superintendent
Mesa Public Schools

EVALUATING THE EFFECTS OF IMPLEMENTING

A DIFFERENTIATED TEACHING STAFF: PROBLEMS AND ISSUES

A Tentative Position Paper for Use
in Project Evaluation

by

Fenwick W. English
Larry E. Frase
Raymond G. Melton

November, 1971

TABLE OF CONTENTS

PURPOSE

Many teachers have asked the project staff how the program at the differentiated staffing pilot schools will be evaluated. This paper is an attempt to establish the framework for the discussion of a design. To some extent, we tried to keep it simple and to the point, but found the usual problem of trying to state some concepts in general terms and thereby lost their special significance. In those cases, we reverted to the more technical vocabulary of the researcher.

Basically, we tried to deal with the issues of what we are trying to assess, problems and issues relating to that assessment, and what kinds of data we would need to gather in order to answer the questions asked. We were not completely pleased with the results, neither from the viewpoint of creating an easily understandable explanation, nor of dealing with all of the highly technical and complicated problems involved. Perhaps, those two objectives are simply not compatible.

We would appreciate comments and suggestions from either perspective, since we are of the opinion that criticism is the impetus for growth.

Fenwick W. English
Larry E. Frase
Raymond G. Melton

Background

Since the inception of the Arizona-Mesa Differentiated Staffing Project in June of 1970, the project staff was charged with the responsibility of evaluating the effects of the changes brought about as a result of implementation in three pilot schools: Fremont Junior High, Lincoln and Holmes Elementary Schools. Accepting this fact in the spirit intended, it was imperative that the school district and the sponsoring agencies, notably the Mesa Education Association, obtain some data regarding how successful the project had been.

The question of what is meant by "successful" can be approached from several different directions. First, if one examines the objectives for which the project received multi-year funding, it becomes apparent that with the mere implementation of a differentiated staffing pattern that the project may be judged "successful." These objectives were:

(1) The creation of site specific differentiated teaching staffs at the pilot schools which meet both the criteria of the schools, the needs of the learners involved, and the U.S. Office of Education guidelines;

(2) A system of internal performance contracting as the accompanying pay vehicle for the models of differentiated staffing developed at the pilot schools.[1]

However, with this interpretation, differentiated staffing was conceptualized as an end in itself, and implementation satisfies

1 - Mesa Education Association, "Differentiated Staffing, the Mesa Approach," Clarence E. Huber (ed.) Mesa Public Schools, Mesa, Arizona, August, 1971, 24 pp. Offset.

all the requirements for completion of the project. Many differentiated staffing directors in other projects have stopped here saying that that was all that was required.

This, however, does not satisfy the local need for answers to questions like "should the differentiated staffing training program be expanded to include more schools," and "does staff differentiation enhance education for the learner?" Answers to these questions forced the project staff into a more careful analysis of the effects of the differentiated staffing project over the implementation period.

In order to answer these more far-reaching questions, the staff turned to experimental design and found that research in an on-going social system was extremely difficult (some say impossible) simply because the ability to control the internal and external variables were beyond the capacity to regulate. Furthermore, the "treatments" to the pilot schools were not standardized, and they were compounded by the fact that they occurred almost simultaneously, and therefore were impossible to separate and assess in isolation. Furthermore, it was impossible to separate the pilot schools in terms of previous "treatments" from the larger school system. No continua of development were available upon which to locate a "ground zero" or "baseline" prior to the initiation of the differentiated staffing project.

The project staff was aware that a staffing pattern (a methods-means for deploying human talent in any organization) was insensitive to the idiosyncratic dimension of teacher behavior where specific instructional sequences may be located which actually might make a

difference with groups of children. Hanley refers to this dimension as "techniques for progressive changes in the topography of the classroom response under study (which may) have been classified as reinforcement, punishment, extinction and shaping procedures."[2]

This awareness served to force the analysis to a level of abstraction which largely ignored the idiosyncratic dimension where the peculiar behavior of a given teacher in a given situation with a given set of pupils and a given range of learning objectives was analyzed and compared. At this point in time, the admonition by Hanley appears appropriate.

> "....early research must always concentrate all of its efforts in order to demonstrate that a phenomenon can be produced, and subsequent research must examine more closely the factors that produce the phenomenon."[3]

Identifying A Design

There are a growing number of educational evaluators who have advocated a movement away from experimental design in evaluative innovations in ongoing social systems. For example, Provus[4] has developed an evaluation model for ongoing programs which is based upon precise definitions of output, processes and inputs. However, too much time and program activity had passed for this model to be used with the Mesa project. A similar position has been taken by

---

2 - Edward M. Hanley, "Review of Research Involving Applied Behavior Analysis in the Classroom," Review of Educational Research, 40:5 (December, 1970) p. 617

3 - Ibid. p. 611

4 - Malcolm Provus, "Evaluation of Ongoing Programs in the Public School Systems," Educational Evaluation: New Roles, New Means (Chicago: University of Chicago Press, 68th Yearbook, NSSE, 1960) pp. 222-242.

Weiss and Rein[5] in attempting to evaluate "broad-aim" programs. These programs are those "which hope to achieve nonspecific forms of change-for-the-better, and which also, because of their ambition and magnitude, involve unstandardized large-scale interventions and are evaluated in only a few sites."[6] Technical problems with the application of experimental designs for "broad-aim" programs are: (1) there is difficulty in selecting satisfactory criteria, (2) the situation is essentially uncontrolled; (3) treatments are not standardized; and (4) the experimental design is limited in the information it can produce.

The problem with selecting criteria is that precise changes desired may not be very specific to assess so that those responsible do not know exactly how the sought after change will occur. Furthermore, "predetermined criteria, whether only a few or many are employed, present another problem: programs representing large inputs of resources are likely to produce unanticipated consequences whose importance may rival, if not outweigh, the intended ends."[6] This phenomena has already happened with the Mesa project, where the development of concrete procedures for internal performance contracting were developed at least one year sooner than expected, due to some unanticipated consequences of developing a vehicle by which actual training funds could be dispersed for training activities utilizing performance contracting for teacher criterion-referenced training objectives.

5 - Robert S. Weiss and Martin Rein, "The Evaluation of Broad-Aim Programs: Experimental Design, Its Difficulties and An Alternative," Administrative Science Quarterly 15:1 (March, 1970) pp. 97-109.

6 - Ibid. p. 102.

The situation within the Mesa Public Schools was essentially uncontrolled. Schools for the staff differentiation project were not selected on some random basis, they applied, i.e., they "volunteered." While control schools were solicited at a later date, they could withdraw, i.e., they too were volunteers. The best efforts of the differentiated staffing project staff to select control schools without any known contamination of system wide treatments were, therefore, frustrated. At least one major federal project in another area related to staff development had been introduced system-wide at the time control schools were being solicited.[7] Furthermore, training activities were not uniform, except for approximately five weeks of the summer training strands of the 1970-71 school year. The remainder of the differentiated staffing training was site-specific based upon a needs assessment design peculiar to each differentiated staffing school.

Weiss and Rein advocate, among other techniques, the use of the case study as a methods-means for determining the efficacy of a social innovation.

> "Research cannot merely document that the program failed and go on to study a modification of the program; it must identify the causes of failure. In this way the experience can become a basis for designing more effective programs."[8]

Weiss and Rein posit that research activities on "broad-aim" programs should not attempt to ask "did it work," but instead "what happened?" For this reason, the evaluation design of the differentiated staffing project will employ the technique of a case study.

---

7 - The Career Education Project

8 - Ibid. p. 103

However, Campbell[9] for one, is opposed to abandoning the rigors of the experimental design completely. He advocates "quasi-experimental" designs which while not true experiments, nonetheless, have some advantages over the lack of controls altogether.[10]

Even if the experimental situation is not completely under the control of the experimentors, as long as they are aware of what specific variables they fail to control, progress can be made. It is possible to confirm a treatment effect by examining all of the rival theories of why an effect was recorded. Accepting a treatment as the cause of an effect means eliminating or reducing rival explanations of why an effect may have occurred. Even with stringent controls lacking, perhaps several rival hypotheses may be elminiated or weakened with those variables which were controlled. For this reason, the differentiated staffing staff has chosen to employ a quasi-experimental design to formulate tentative hypotheses regarding the effects of training at the differentiated staffing pilot schools. Campbell's criticism of abandoning the experimental method completely appears to be justified.

> "In the present political climate, reformers and administrators achieve their precarious permission to innovate by overpromising the certain efficacy of their new programs. This traps them so that they cannot afford to risk learning that the new programs were not effective."[11]

---

9 - Donald T. Campbell, "Considering the Case Against Experimental Evaluations of Social Innovations," Administrative Science Quarterly, 15:1 (March, 1970) pp. 110-113.

10 - Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching. (Chicago: Rand McNally, 1963) pp. 171-247.

11 - Donald T. Campbell, op. cit.p. 111

Campbell and Stanley[12] discuss factors jeopardizing internal and external validity of experimental treatments. They identify eight classes of "extraneous variables" which might produce confounding effects along with the "treatment" which damages the internal validity of the experiment. Of these five of eight would apply to the internal situation of the differentiated staffing project in the Mesa Public Schools.

Factors Jeopardizing Internal Validity

History: Events between treatments may confound the isolation and measurement of the treatment from some historical event. In the Mesa Public Schools such events may have been the introduction of an OEO "Incentives Only" program, teacher effectiveness training, etc. In addition, visitation to the pilot schools by outside professionsl could have produced a Hawthorne effect with the staff and student body which could be indistinguishable from the actual training program.

Maturation: Since the project spanned a period of two years, maturation on the part of the participants was evident in gaining new insights, changing responses, becoming fatigued, etc.

Selection biases: Schools were not selected randomly. There is evidence to indicate that at least two of the three schools which volunteered were previously considered to be "innovative: and had undergone staff changes

12 - Donald T. Campbell and Julian C. Stanley, Op. cit.

and participated in other programs of staff development. Teachers in these schools were recruited especially because they desired to become part of the on-going "innovation." In addition, students who attend these schools could not be randomly selected across the school district. Therefore, addressing "gain" scores of students is complicated by neighborhood characteristics, home environments, scholastic aptitudes and classroom inter-action styles.[13] There may be some pattern at the three schools which is atypical or which if not present in an "average school" would account for a hypothesized gain in scores on tests.

Experimental Mortality: Two of the schools involved with staff differentiation experienced a loss of personnel which may have lessened the effect of the "treatment." Fremont experiences a 16% loss of teachers who originally began the program, Holmes 37% and Lincoln 6%. The differentiated staffing staff could not control marriage, pregnancy, retirement, transfer or promotion at these schools. This situation was compounded by the selection biases of the pilot schools in "replacing" "lost" personnel.

---

13 - Barak Rosenshine, "The Stability of Teacher Effects Upon Student Achievement," Review of Educational Research, 40:5 (December, 1970) p. 651.

Factors Jeopardizing External Validity

Campbell and Stanley also address themselves to factors which appear to confound and reduce the external validity of an experiment. External validity is concerned chiefly with conditions which lead to experimental generalizability. Those which apply to the Mesa project are as follows:

(1) Selection biases previously indicated under a discussion of internal validity;

(2) Multiple-treatment interference due to an inability to "erase" previous innovations and training experiences over an extended period of time, thus compounding the multiple effects of both treatment, maturation, and selection bias.

(3) The combination of selection of teaching staff, going with both volunteer experimental and "control" schools rather than random selection, means that both experimental and control schools may have been atypical in the first place. In order to engage in the innovation it may have required the schools to be more open, more positive towards change, possess higher morale, be more professionally competent and secure than schools selected "at random."

TENTATIVE RESEARCH DESIGN

QUESTION: Did the differentiated staffing training program make any significant difference with student achievement or pupil attitude growth?

HYPOTHESIS: No significant statistical difference between differentiated staffing and non differentiated staffing schools. Alpha = .05

DATA

1. Pupil achievement test scores on standardized achievement test;
2. Pupil attitude scores on pre and post test.

Independent Variables

1. DS/non DS schools; (two levels);
2. Sex of the responding teacher (two levels;
3. Age of the responding teacher (four levels);
4. Class size of the responding teacher (four levels);
5. Attitude of the responding teacher (MTAI percentiles);
6. School climate OCDQ (six levels);
7. Degree of instructional individualization (IOTA);
8. Degree of humanization (IOTA);

Statistical Treatment

1. Pearson Product r/ correlation table;
2. Multivariate analysis of variance (MANOVA);

Confounding/Unaccounted Effects

1. The degree of skill possessed by the staffs prior to specific staff training as part of the DS project;
2. The degree of skills possessed by the control school staffs and possible "ripple" effects in skill acquisition as a teacher in the Mesa Public Schools; (being affected by other system inputs);
3. Lack of random student selection. Peculiar neighborhood characteristics of the student body, aptitudes, etc.;
4. Replacement of teachers due to turnover was not random; selection biases both in the initial choice of the schools and later in teacher replacements;

Discussion

To some extent the previous "experience" indicator in a teacher questionnaire from the teachers would indicate the amount of training on a formal level in the teacher's professional background. It may be assumed that the "ripple" effect would influence all schools more or less the same since the DS schools were not isolated from the total system. Problems of randomization cannot be overcome.

QUESTION: Did those schools which received the differentiated staffing training inputs possess more "open" school climates and better faculty morale than those which did not receive those training inputs?

HYPOTHESIS: No significant statistical difference between DS and non DS schools.
Alpha = .05

DATA

1. Likert scales which are translated into one of six climate categories;
2. Inferred status of faculty morale from OCDQ categorization;

Dependent Variables

1. OCDQ classifications.

Independent Variables

1. Sex of the responding teacher (two levels);
2. Age of the responding teacher (four levels);
3. Professional experience of the responding teacher (three levels);
4. MTAI percentile scores;
5. Pilot/control schools (two levels);
6. Class load (four levels);

Confounding/Unaccounted Variables

1. Without pre-test data when the project began, it may be argued that the schools which volunteered for both project participation and those which volunteered for control school status would be already more "open" than if they had been chosen at random. If a relationship is found to be statistically significant in terms of the training program, this fact must be heavily considered;

2. Loss of personnel at the pilot schools may decrease the effects of the training inputs.

Statistical Treatment

Same as with the first question.

Discussion

Accepting the null hypothesis of no significant difference may in fact not be a true indicator of the effects of the training on faculty morale or school climate. This can be determined by an examination of precisely what types of climate are found as a result of the testing.

QUESTION: Did those schools receiving DS training inputs possess more pupil centered staff attitudes and greater instructional individualization practices and program humanization than those without such inputs?

HYPOTHESIS: No significant statistical difference between DS and non DS schools.
Alpha = .05

DATA

1. MTAI teacher scores (.05 level or better by comparison);
2. IOTA scaled observations on degree of individualization currently a part of the program (.05 level or better);
3. IOTA scaled observations of the degree of humanization currently a part of the program (.05 level or better);

Dependent Variables

1. MTAI scores;
2. IOTA individualization scores;
3. IOTA humanization scores.

Independent Variables

1. Pilot/control schools (two levels);
2. Sex of the responding teacher (two levels);
3. Age of the responding teacher (four levels);
4. OCDQ indicator (six levels);
5. Class size of responding teacher (four levels);
6. Experience of the responding teacher (three levels)

Statistical Treatment

Same as other two questions.

Confounding/Unaccounted Variables

1. In the absence of baseline data prior to treatment, it may be hypothesized that the DS schools already possessed greater positive attitudes towards children and were practicing both individualization and humanization prior to the training program. Did the DS inputs begin a trend or accelerate a trend?

2. Can it be assumed that the control schools were without any treatments? What effect did other federal programs or school initiated in-service training produce the same results?

3. To what extent were trends accelerated by the recruitment of teachers with special proclivities for the type of program already functioning?

Discussion

The lack of baseline data prior to treatment is a definite handicap plus the lack of randomization of selection, replacement, and lack of control of other factors. To some extent this may be subjectively judged by constructing a teacher questionnaire to be administered at both schools which attempts to assess where the teachers thought they were prior to the training program. Such a questionnaire would assess what teachers at the control schools had acquired by the way of in-service or university based training would be revealed.

The situation may reveal that there still is a statistically significant relationship between the DS training inputs and the development of better teacher attitudes towards children in which case an alternative hypothesis might be that the teachers already possessed these attitudes and that those that didn't dropped out because of some aspect of the training program thus leaving by default (and

complicated by non-random replacement) teachers with more positive attitudes. The reverse may also be true. Other factors which may or may not relate to the training program may be responsible for a significant relationship.

QUESTION: Have the differentiated staff project inputs (training, travel, experience) enabled the teachers of those schools to more adequately respond to the tasks implied or stated in the RFP than those at the control schools?

HYPOTHESIS: No significant difference between DS and non-DS schools: significant defined as a 20% difference or greater.

DATA:

Data will be gathered on a case study instrument (questionnaire which assesses situational conditions, etc. which pertained and were involved) followed by interviews of all teachers at the DS schools.

Specifically, the data gathered would be:

1. Frequency of the occurrence of particular types of problems in responding to the RFP;
2. Frequency or amount of time spent in responding to the RFP (it may be hypothesized that less time means that the schools were better organized and knew how to handle the requirements of the RFP);
3. A rating of the school-produced RFP by the Project Evaluation Panel on criteria to be developed as also influenced by problem frequency and time spent.

Dependent Variables

1. Frequency counts on problems and time spent;
2. Rating of RFP's;
3. Anecdotal comments; (non-statistical);

Independent Variables

1. Pilot/control schools (two levels).

Statistical Treatment

Percentage bar-graphs computed of types and frequency of roles differentiated; problems encountered, etc.

Confounding/Unaccounted for Variables

1. The clarity of the tasks may make the skills required rather obvious, i.e., perhaps it is more of a measurement of the specificity and clarity of the RFP than actual skill input derived via training; or this may be a confounding variable with training.

2. To some extent the observers are limited to the memory recall of the participants; thus, limited to the extent pleasant things are recalled, unpleasant things forgotten;
3. Time spent may be more of an indication of the size of the group than of the tasks required.

Discussion

Very few other total school staffs in the district have had experience with the RFP procedure. To a limited extent, it has been similar with the CID district program to which other schools have applied. Some of the control schools may have had some experience in this situation or some of the teachers at these schools.

QUESTION: Did the schools with the differentiated staffing training inputs arrive at a greater degree of on-site role differentiation than those without such inputs during the period of time in which they were responding to an RFP? Were such role differentiations functional in terms of meeting actual RFP requirements for teacher/pupil products and pupil achievement?

HYPOTHESIS: No significant difference between DS and non-DS schools. Significant defined as 20% difference or greater.

DATA: Data gathered in the following manner:

1. Teacher questionnaires which assess the degree of task differentiation;
2. #1 based upon on-site observation by trained observational teams; (creation of frequency categories);
3. Anecdotal data on a cross-reference basis determining if role differentiation was function in terms of tasks necessary to develop an acceptable (to the Evaluation Panel) bid based upon specifications contained in the RFP.

Dependent Variables

1. Frequency counts of types of role differentiation which actually occurred during the bid period;
2. Matching of types of differentiation to evaluation of bid (numerical rating) and to number of students who met the terminal objectives required in the RFP.

Independent Variables

1. Pilot/control schools (two levels);

Discussion

Besides problems of rater reliability and instrument reliability, all of the issued with the previous questions and the case study procedure would apply.

PRESENT INSTRUMENT RELIABILITY/VALIDITY

	Instrument	Reliability	Validity
1.	Minnesota Teacher Attitude Inventory (MTAI)	.92	.60
2.	OCDQ (Organizational Climate Description Questionnaire	.62	Unknown
3.	IOTA - 8 scales "humanization"	.89	content
4.	IOTA - 4 scales "individualization"	.89	content
5.	Achievement Test-Grade 9 Mathematics (new Met '70)	.90	content
	Reaching (new Met '70)	.92	content
6.	Standard Achievement Test Math - grades 1-8 (new Met '70)	.86-.92	content
	Reading - grades 1-8 (new Met '70)	.87-.92	content
7.	Dutton's Attitude Inventory Towards Math	.94	content
8.	Inventory for pupils towards teachers	.73	not established
9.	Inventory for pupils towards education	.80	not established

DEFINITION OF TERMS

1. "Differentiated staffing" - a process whereby specific teaching skills are identified in relationship to meeting specific pupil needs.

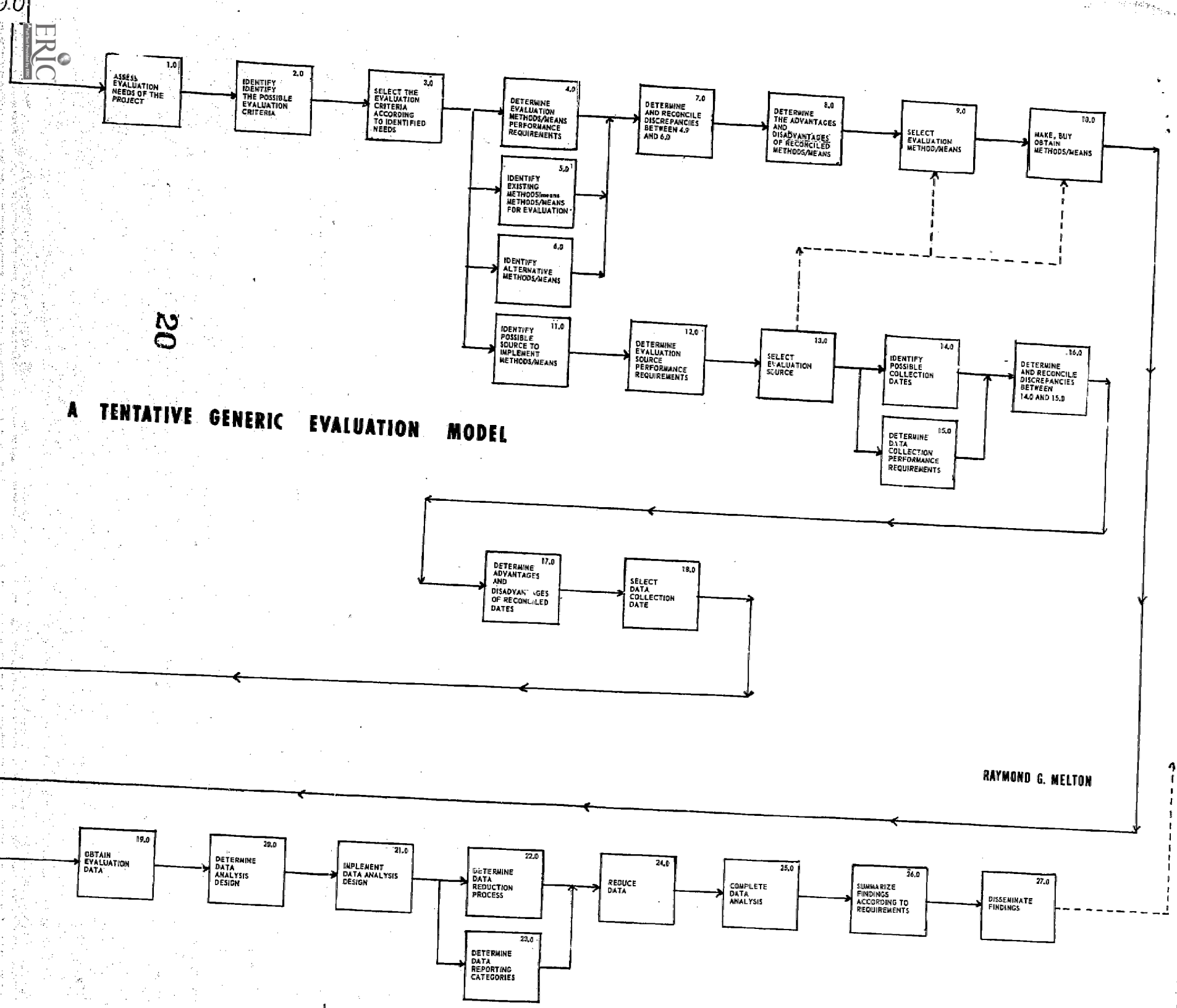
2. "RFP" - Stands for "Request For Proposal." It represents the specifications for schools to submit estimates of costs of how much resources will be necessary to meet the requirements spelled out in the RFP.

3. "Training." Stands for a variety of staff inputs, some involving the presence of a trainer and trainee, some with sets of instructions and no trainer. All of the activities below have been part of the "training inputs."

   (a) inquiry training;
   (b) teacher effectiveness training;
   (c) group dynamics;
   (d) media application;
   (e) flexible scheduling;
   (f) on-site observation of the utilization of "open space;"
   (g) PERT workshop;
   (h) system analysis;
   (i) "Bidding" workshop and mini-work sessions;
   (j) various facets of evaluation;
   (k) experimental curriculum development;
   (l) assessment instrument development;
   (m) needs assessment model development;
   (n) needs assessment;
   (o) affective curriculum and material development;
   (p) development of educational philosophy;
   (q) facility evaluation;
   (r) development of job descriptions and pay scales;
   (s) development of various time/space/staffing models;
   (t) preliminary development of PPBS techniques;
   (u) primary task analysis;

4. "Bid." A statement by a school in response to an RFP which lists the activities and objectives it feels are required to meet the specifications of the RFP. It is usually accompanied by a budget as to how much such activities and objectives will cost. Cost is usually represented as teacher training time.

A TENTATIVE GENERIC EVALUATION MODEL



RAYMOND G. MELTON